

6/23/59

Dr. Joshua Lederberg
Dept. of Genetics
School of Medicine
Stanford University
Stanford, California

Dear Dr. Lederberg:

I was most happy to receive your letter of June 18th. I hope that you had a pleasant trip. And I was most interested to find out that your first letter was not stimulated through Godon Allen. Incidentally I haven't heard from him lately or the Amer. Soc. of Human Genetics.

It is reassuring to find that even you, as a supporter of the Citation Index idea agree that advance work should be done to work out the bugs. This was precisely what I had in mind when I submitted my original NSF proposal. I think this is evident upon reading the proposal which is now enclosed along with some other papers and correspondence.

I can't agree that in this instance the reason for the turn down was the financial condition of NSF. I am certain that many other proposals in other divisions of NSF get turned down because there isn't enough money to go around, but in the Office of Scientific Information they go around pleading that nobody wants to do research in documentation and always have. What they mean is that nobody wants to do the kind of research they want. In fact, it is a crime that almost all of the money they give out is for projects which in a certain sense they originate themselves. They just signed a contract with Itek Corp. for \$140,000. I'm still not sure what it is for. I happened to find out by seeing a Stock Market Prospectus issued by this firm. They also give out money for ~~xxx~~ "popular causes" like translation of Russian stuff--regardless of its scientific value. You can't imagine how frustrating it has been in the past five years (or maybe you can) to have had at the helm of scientific documentation activities in NSF a woman who was neither a scientist or an information specialist, but just a good secretary (a Spanish major) who worked her way up by taking good notes at meetings and preparing reports for her bosses. I would never say this publicly, but that is the absolute truth. I tried for five years to get some kind of support so I wouldn't have to go "commercial" but it was a losing battle. I even got myself temporarily affiliated with the Univ. of Pa. ICR and the Franklin Inst. and couldn't make a dent.

You are probably absolutely right about ~~a~~ going to other agencies. I think I should have tried the AFRDC long ago, but just wasn't sophisticated enough. Actually, I did try NIH (as the enclosed will show) but even then I should have submitted the proposal through the regular grants office as I think I will in future. I also tried ONR but they turned me down too--even though they were sympathetic. I never approached AEC, but if you knew some of the dolts in their Information set-up you'd soon agree that might also have been fruitless. I regret to say that a few

of these people have now gone over to NSF. They still can't understand why CURRENT CONTENTS is so popular. What people really need is more abstracts. By the time they get the entire literature abstracted--selectively or otherwise CURRENT CONTENTS will be, I think, making me, at last, a nice income. I recently suggested a Space Sciences edition of CURRENT CONTENTS and the NSF-NASA (former AEC boys) still can't see it. Consequently I have given up for the moment. When we have the capital we will do it on our own and I am sure we will make a handsome profit. It is a tremendous field with most inadequate information services.

You are right, I think, in your comments about the applicability of Citation Indexes to biology and medicine rather than chemistry, although I have Citation Indexes as an extremely cheap method of bringing together papers on a specific compound. In our steroid coding project I try to use this principle all the time. However, you are right that CA does a fairly good, though belated job, and it is a tough battle to get them to change.

I have always stressed that Citation Indexes are no substitute for subject indexes. This is true of the legal literature too. First one uses the "digest" to find an interesting case or two and then uses the citator to locate the cases that have subsequently emanated from these.

You are so right about the manpower aspect of indexing. CA boasts that it will catch up in indexing by 1962 at which time they will be back at their old schedule of only being six months late with the yearly index.

It is a funny coincidence that you should mention the cost of a key punch operator in Italy. I've been corresponding with a fellow in the FAO in Rome who has been doing a sort of Citation Index on cards (3X5) in the field of fisheries biology). We are discussing the possibilities that his staff would do the leg work on our project--or at least on that portion he could justify. The costs would be about 50% less than over here. And as you say they can handle the foreign languages easier. By the way, even clerks with imagination can't handle Japanese citations! We'll need some clerks with a knowledge of specific foreign alphabets like Japanese, Russian, etc. Russian doesn't really bother me as you can train a girl to transliterate in about one hour.

I've taken you up on your offer to read my proposal--it is now enclosed.

I am sorry but you are trying to give the Patent Office people credit for more intelligence than they have. You don't know how backward they are. It is such a tradition bound organization that even their approach to machines, which they are investigating, is completely archaic. I suggest you meet their Dir. of Research some day if you want to be convinced. They did not reject the Citation Index on the grounds you suggest--it was purely on the grounds that they didn't think it was worth the effort. You can't find out whether a new patent has subsequently issues on one which you are interested--you frequently find useful references to earlier patents in the patent, but these are not references in the usual sense. See the enclosed paper. The crime of the Patent Office story is even worse as regards their own internal procedures. Not only should published patents be Cited, but the files of rejected patent applications are even more important, because each one contains a wealth of search information that has already been

worked up by an earlier search on the same subject. In other words, if you file an application on an invention and it is rejected as being covered by the prior art--and then I come along next year and file on the same thing--they go through the same damn procedure. There is no simple way for the new examiner to know that such a search has already been done. Since only 50% of all applications result in patents this means that every other application is on old stuff. In addition, of the inventions patented most have more than a half of the claims rejected. And congress wonders why it takes over two years to get patents and sometimes longer.

The one exception to this now is the steroid art in which they are using a very simple punched card coding scheme. I'm enclosing the code sheet we use. (We have a contract to code and screen the literature for new steroid chemicals). However, the real reason they have been able to cut down on search time is not because of the virtues of this code, but because they got away from the old classification system. They would never admit this. Now, instead of going ahead with simple but effective methods such as this one they are playing around with all kinds of fancy ideas that may pay off ten years from now. In the meantime you can wait a long time. I have an application in since last August and I haven't even gotten the first action yet. After I do it will still take a long time to get through with it. You can't imagine how much this stultifies what I've been trying to do with my invention. (I've been working on a selective copying gadget).

Returning to your suggestions on a reasonable experiment for a citation index project--I am grateful for these. Your idea of starting with a review journal is most interesting. Actually it is just the reverse kind of thinking I once applied in a paper in which I suggested that we use review articles as a source of index entries. However, I never thought of using the Reviews as the starting point for a citation index chain--and now that you mention it I think I can see the logic--I guess I didn't fully appreciate how much review papers are cited today--I know that review papers are highly valued by most scientists--but I didn't know they were cited in the way you mention. Perhaps this has to do with the definition of a review paper. If you have a copy of the Review which you wrote handy I'd like to look it over. Could you mention a few points in it that you are particularly interested in--what subsequent ramifications might you expect or do you already know have developed?

Of course, in suggesting the kind of test that you did you are placing us in the position of comparing the Citation Index with the effectiveness of the conventional indexes. In the enclosed paper I did this for patents.

I don't know exactly what it would cost to conduct the experiment that you have in mind. It wouldn't be cheap as it would involve a lot of leg work--correspondence and testing and psychological factors, etc.. However, let me give it some thought and find out whether I could produce the basic corpus of references needed without too much trouble. Whoa! You said all the journals we cover in Current Contents. 450 journals X 12 issues per year X 15 articles per issue X 6 for the six year since 1952. You can scan journals pretty quick. In looking for steroid articles we go through the journals page by page. When we don't find anything it goes very fast. Let's figure 10 minutes per issue or half a minute per article. That might involve 250,000 minutes or 4,000 hours--2 man years of work. However, for the experiment you have in mind I think the inter-disciplinary approach is not as important since

you are trying to compare what you find with what you will get out of Biol. Abs. or the Current List. For this reason I think you could easily cut down the amount to be scanned by at least a factor of 80%. 800 man hours is not so bad.

Were you thinking of this experiment as a means of convincing people further of the value of Citation Indexes as compared with ~~xxxxx~~ conventional indexes. In the project I proposed to NSF we could easily have obtained the data you want. Perhaps in rewriting the proposal we can incorporate this as a specific experiment to be done along with others.

I will take up your suggestion about talking to Dr. Koprowski who is an ardent user of Current Contents. Incidentally, I am on very good relations with the people at Biological Abstracts. The Director, Miles Conrad, is a good friend, but I know that he didn't see the point of spending \$30,000 on research on Citation Indexes. He was one of the referees.

I think that I anticipated you on the idea of getting the NIH Div. of Research Grants interested. Their former librarian, Scott Adams, tried to get them interested but nothing came out of it.

None of your suggestions are inane--and certainly not obvious. I would be more than glad to have an opinion from someone like Dr. Jean Duncan. It would take much more time to explain how a computer outfit could use citation indexing as part of a linguistic approach to analysis of documents--but that is really a rough one. (I am doing some graduate course work in Linguistics at Penn. and have been giving much thought to using this principle for mechanical analysis of documents. I believe they will have to come to it ultimately as the primary shortcoming of all approaches I have read about it that they treat each document as a separate entity--whereas each document must be treated, even linguistically, in the relationship it holds with related information in other documents.) However, I think punched cards or their equivalent would really be sufficient for a long time. When our volume of cards really mounts up then more sophisticated methods may be in order.

Well, if my last letter left you in a state of shock, this one ought to leave you in a coma. My only regret now is that it is too late to send in a new application to NIH in time for the July 1 deadline--I am almost tempted to try it anyhow. But I just finished writing an application for a grant for Current Contents. NIH asked me to submit one so that we can reduce the price for individuals. I don't know if I have the pep for another one in the same week.

With best wishes to you and with anticipation of your reply, I am,

Sincerely yours,



Eugene Garfield